DOCUMENT RESUME

ED 176 188 CG 013 791

TITLE	Evaluation Issues.
INSTITUTION	Center for Action Research, Inc., Boulder, Colo.; Little (Arthur D.), Inc., Washington, D.C.
SPONS AGENCY	Law Enforcement Assistance Administration (Dept. of Justice), Washington, D.C.; Office of Juvenile Justice and Delinquent Prevention (Dept. of Justice), Washington, D.C.
PUB DATE	Jun 78
CONTRACT	J-LEAA-013-77
NOTE	22p.
EDRS PRICE	MF01/PC01 Plus Postage.
DESCRIPTORS	Counselor Evaluation; *Data Analysis; Delinquent Rehabilitation; *Delinquents; *Evaluation; Juvenile Courts; Profile Evaluation; *Program Evaluation; *Research

ABSTRACT

The purpose of this paper is to encourage the reader to distinguish between evaluation and other information-gathering procedures, to distinguish between measures of efficiency and effectiveness, and to make informed decisions about when to use evaluation rather than another procedure for producing information about a project. Material is divided into types of information-gathering activities, types of evaluative research, preliminary questions preceding selection of type of evaluation, suggestions for further resources relative to the technical aspects of evaluative research and an outline of tasks involved. Although specifically written for judicial programs for juveniles, these step-by-step procedures can be applied to other evaluation efforts. (Author/BEF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED176188

EVALUATION ISSUES

Prepared For

The Office of Juvenile Justice
and Delinquency Prevention

The Law Enforcement Assistance Administration

U.S. Department of Justice

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

CG013791

This project was supported by Contract Number J-LEAA-013-77, awarded to Arthur D. Little, Inc., Washington, D.C., and the Center for Action Research, Inc., Boulder, Colorado, by the Office of Juvenile Justice and Delinquency Prevention, Law Enforcement Assistance Administration, U.S. Department of Justice. Points of view or opinions stated in this document are those of the authors and do not necessarily represent the official position or policies of the U.S. Department of Justice.

JUNE, 1978

TABLE OF CONTENTS

	Page
INTRODUCTION	1
TYPES OF INFORMATION-GATHERING	2
TYPES OF EVALUATIVE RESEARCH	3
Process Evaluation	4
Impact Evaluation	4
PRELIMINARY STEPS TO EVALUATION	6
STEP 1: Will the findings be used?	6
STEP 2: Is the project evaluable?	6
STEP 3: Who can do this work?	7
TECHNICAL STEPS	9
Recommended Resources	9
Technical Tasks	10
TASK 1: Formulating the Question	10
TASK 2: Designing Instruments or Techniques for Measurement	10
TASK 3: Designing the Study; Common Models for Evaluation Studies	11
TASK 4: Data Collection	12
TASK 5: Utilization of Results	12
APPENDIX A	A-1
REFERENCES	

INTRODUCTION

A good evaluation presents many difficulties. Designing and implementing a good piece of evaluation requires staff cooperation:

- Staff must be willing to accept that rigorous standards will be applied in the evaluation.
- Staff must be prepared for unanticipated setbacks in the research tasks accompanying an evaluation effort.
- Staff must understand that findings will demand careful and thoughtful interpretation.
- Staff must honestly decide whether or not the results warrant the continuation of an evaluation effort.

Evaluation should be considered an appropriate activity worth the effort, rather than a casual project taken up for unclear or unspecifiable reasons. Unless it is regarded as a choice, there are other types of documentation, reporting, and analyses that could be applied to a project and that may be more appropriate.

The purpose of this document is to encourage the reader to distinguish between evaluation and other information-gathering procedures, to distinguish between measures of efficiency and effectiveness, and to make more informed decisions about when to use evaluation rather than some other procedure for producing information about a project. This document is not an attempt to detail how evaluations are designed and implemented. There are good resources already available for this which would help an agency or project staff to determine whether it has the necessary staff, or would suggest what sort of assistance the staff will want to seek outside the organization.

The material is divided into:

- Types of information-gathering activities.
- Types of evaluative research.
- A list of preliminary questions to be asked before selecting evaluation.
- Suggestions for further resources into the technical aspects of evaluative research.
- A brief outline of the types of technical tasks that are involved in evaluation research.

TYPES OF INFORMATION-GATHERING

Three general but distinct procedures are commonly used: assessment, monitoring, and evaluation.

Assessment is establishing the rate or amount of some activity or resource. Assessment of needs is a means of estimating or determining the amount or importance of unmet needs and identifying the resources currently available to meet those needs. Needs assessment sometimes uses existing data or involves collection of new data. It is a method of finding service delivery gaps and substantiating unmet needs in a community and is used to establish priorities for addressing problems.

Monitoring compares a project's plans with what actually happened. It entails collecting specific information on events associated with the operation of a project. In general, a monitoring system obtains data on both the project and its activities, allows for the analysis necessary to determine whether activities are acceptable, and provides for feeding back this information to management. Monitoring activities are associated with reporting systems and cost analysis techniques. Specific information on this function is presented in Monitoring for Criminal Justice Planning Agencies (LEAA, 1974). A more general reference is Monitoring for Government Agencies by John Waller, et al. (The Urban Institute: Washington, D.C. 1976).

Evaluation refers to judging the merit of something by comparing it against some yardstick. Evaluation studies are done to measure the effects of a program or project against the objectives it set out to accomplish and thus aid subsequent decisions about the project's future or structure. In the strictest sense, evaluation studies using systematic methods are designed to verify whether a certain effect occurred and to suggest conclusions about the extent to which this effect can be directly attributed to the project rather than to outside forces.

TYPES OF EVALUATIVE RESEARCH

As evaluation research has become widely used in government agencies and action projects, attempts have been made to distinguish degrees of success or failure. Programs are designed to produce a specific outcome. These results are related to program input and include its purposes, principles, methods, staffing, location, auspices, and numbers and characteristics of persons served. Both input and outcome can be evaluated. (Suchman, 1967), Evaluation Research has come to describe systematic measurement and analysis of the effects of various aspects of program review. Evaluative research is involved when activities are judged against implicit criteria, even if they are not judged against the actual outcome of program activities.

The following are examples of areas to which evaluative research might well be applied:

- Effort. This refers to both quantity and quality of activity. There is the assumption that the specific activity is a valid means of reaching the objectives, so it is measured regardless of output. Monitoring activities usually cover at least quantity of activity.

- Efficiency. The evaluation of alternative methods of achieving objectives in terms of costs involved -- money, time, personnel, and public convenience -- focuses on the efficiency with which procedures are carried out. It is often done as cost benefit or cost effectiveness analysis.

- Operation. The focus here is on detailed questions concerning why a program works or does not. A description of operation would include program attributes, program recipients, program conditions, and details about differences or changes in these and their effect on the success or failure of the program.

- Effectiveness of Performance. This is the measure of results of effort, rather than of effort itself. The key question concerns the extent to which program objectives have actually been attained. Methodological considerations and controlled analysis are needed if the study must also demonstrate that what occurred was a result of project activities.

- Adequacy of Performance. This compares performance to the amount of need: How much of the entire problem has been solved as a result of the program or project? Effectiveness measures impact on a population actually served; adequacy measures impact on a larger population of need.

These five categories combine to form two basic types of evaluative research: measures of the inputs or processes in the project and measures of the outcome or impact of the project. While each provides a distinct type of information, used together they provide a balanced analysis of the program.

Process Evaluation

This is a detailed study of the various parts of the system and how they are linked. Such a study can be accomplished through careful and systematic documentation of staff activities, project attributes, recipients, etc., and the conditions and changes in any of these. Carefully designed reporting systems should provide for all of these things, as well as for adequate information to make estimates, using cost analysis techniques, of the efficiency of the activities and services. This involves more than monitoring project activities and careful documentation of project characteristics. It also requires judging the quality, adequacy, or appropriateness of the procedures and making necessary adjustments. There must be some rationale for analyzing these aspects of a project. For example, keeping records of how many staff hours were spent talking with clients one-to-one and how many hours were spent talking to clients in groups has a reason if the program hopes to find out if one-to-one contact is more or less valuable.

Impact Evaluation

The concern here is with the relationship of project outcomes to stated goals. It is assumed that a problem has been identified, that there is a theory about what will alleviate it, that the project has a goal, and finally, that the program activities will have a specific impact on the target group.

An impact evaluation depends on a statement of the theory of what will alleviate the problem. It determines how the project is set up and allows for measurement of the proposed effect. At a minimum, an impact evaluation will be a study of change in some target group (individuals, a section of a community, a jurisdiction, etc.).

However, it is even more important that the study be designed for conclusions about the extent to which the project activities themselves created the change. For example, to say that a job program for youth reduced their delinquency behavior, it would be necessary to have a "control" or comparison group of similar youth who were not provided jobs. Both the group provided jobs and the control group would be measured by their delinquent behavior before and after the job program. If delinquency were lower for those given jobs than for the control group, it would be more persuasive evidence of the impact of the job program than a before and after measure on only the working group.

Although the sophisticated design and rigorous technique used to link project activity and effect on its target provides the best type of information in deciding about program design and existence, impact evaluation is not always appropriate or possible. It may not be feasible because of limited resources, limited expertise, or inadequate program design. Or it may not be desirable because no one intends to act on that type of knowledge.

Process and impact evaluations are related. For planning purposes, knowing that a project had a certain effect may not be as useful as knowing that it had an effect and why it had that effect. Using the job program example again, it may be found that the impact on delinquency was uneven across the group getting jobs. With adequate information on how the jobs for youth were created and what those jobs entailed, it may be possible to determine that different types of jobs were viewed differently and thus had different impacts on reducing delinquency. For example, jobs that the employer created and financed may have been viewed more favorably by youth than those jobs created by employers simply because the job program was providing the salary with no real commitment made by the employer.

The LEAA guidelines call for State Planning Agencies to make "intensive evaluations" when planning warrants it. An "intensive evaluation" implies a study including both process and impact evaluation which would facilitate adjustment of project activity or transfer of project design to others. Both process and impact must be described carefully enough so that others wishing to design a similar program for a similar impact on youth can provide similar staffing, service facilities and clients, or, if necessary, adjust aspects of project activities to produce an even more favorable impact.

The information gained from monitoring and documenting the activities of a project is frequently all that is required. A descriptive report of what was done at the end of a specified period may be sufficient. Even descriptive reports, however, require thoughtful consideration and analysis in order to be useful for purposes other than that of establishing compliance. There is often room for improvement in the way activities are carried out, especially in the way they are planned for and incorporated into the project from the beginning.

However, project staff and others may decide that an evaluation of the project, particularly an evaluation of the project's impact, is what is wanted. They want to go beyond monitoring ongoing activities and determine if the project activities are producing the projected outcome, and if this outcome is related to activities and not to other forces. The following section discusses preliminary steps in deciding whether or not such evaluation actually can and should be undertaken. There are three crucial questions to consider carefully before commitment to an evaluation effort: (1) Will the findings be used? (2) Is the project evaluable? (3) Who can do this work?

PRELIMINARY STEPS TO EVALUATION

STEP 1: Will the findings be used?

When considering evaluation, first ascertain whether there is a sincere interest in or need for the findings, even if they are not entirely favorable. Who wants to use the results and for what purpose? Careful consideration should help to determine whether to go to the trouble and expense of doing an evaluation and should guide the decision as to whether it will be a process evaluation or an impact evaluation. Evaluation should not be performed on the basis of such reasons as "it is required" or "it ought to be done because others do it." It should be done because the information is needed for decision making and will be taken into account in future planning.

Another factor in considering whether an evaluation study actually would be used is that of the distinction between summative evaluations and formative evaluations (Scriven, 1967).

- Summative evaluations aid decisions on whether to keep or stop a project.

- Formative evaluations aid decisions on whether changes are needed in the way the program is structured or in its operations and, if so, whether or not to make these changes.

Agencies or communities with an experimental approach will most often be interested in formative evaluations: their interest will lie in improving programs, since that is the more efficient way to plan and to develop a program. Formative evaluation is appropriate for feeding useful information back into the planning process.

"Summative" and "formative" concern more the use of the evaluation than how it is conducted; but, if it is clear that formative evaluation is needed, process evaluation with impact evaluation must be included. It is not possible to change a program's operation to improve outcomes if the original processes are not understood and have not been analyzed with respect to their results.

STEP 2: Is the project evaluable?

Once it is decided that an evaluation is needed, the next step is to determine whether the project is evaluable as currently designed and operating. The key conditions are:

- Ability to identify the project's objectives. What is this project trying to accomplish? Too often this question cannot be answered clearly and precisely. Objectives must be stated in

terms that are clear, specific, and measurable. The key assumptions of the program must be stated in a form which can be tested objectively. That is, not only must the outcome be definable, but also the process used to achieve it must be specifiable.

- Ability to identify the criteria for data. The final criteria must be specified. How is the outcome going to be measured? For example, will it be delinquency rates or will it be self-reported delinquency among the clientele?

- Ability to identify the relevant groups that the evaluation can use to measure impacts. There must be a specifiable clientele group or population where the change is expected to occur directly.

In determining whether the project can be evaluated, it may become evident that there are multiple objectives or multiple definitions among the staff and that there are both short-term and long-term objectives. Evaluator and evaluatee must select among the objectives one or two which are more important or more interesting; they must be able to agree about definitions; and they must be able to decide whether to measure short-term or long-term effects, or both. If these things cannot be agreed upon, then the use of the findings of the study is jeopardized.

STEP 3: Who can do this work?

Most of the steps in an evaluation do not require highly specialized professional training as much as they require staff with a commitment to being rigorous, systematic, and objective. Often the agency overseeing evaluations or a project carrying out its own evaluation will have staff capable of undertaking many parts of evaluative research. However, it may not mean that they also have time and this needs to be taken into consideration.

Even with capable staff, it is still important to have someone with expertise and experience in this type of research to direct the technical aspects of evaluation and to design the study and the measurement instruments to be used. There may be such a person on the staff. If not, there are professional evaluators within most state government systems or available for hire from private firms in the community. However, even selecting someone to hire takes some knowledge and understanding of the process. Agencies intending to do much evaluation or to oversee evaluations as an ongoing activity should acquire staff with this capability. A checklist of what to look for in an evaluation plan is Appendix A.

Another alternative is to limit the types of evaluation to those which the available staff can handle. Attempting a sophisticated and complex design without proper understanding of the techniques or skills to do the necessary statistical analysis and data interpretation would waste staff time and project resources.

Therefore, in deciding on staffing needs, consider three things:

- What sort of evaluations will be conducted (e.g., process or impact, short-term or long-term, etc.) and how many will be conducted?
- What funds are available for the evaluation activities?
- What parts of the evaluative research could be done by in-house staff and what parts by outside contractors or consultants?

Somewhat more in-depth discussions of the management of resources for evaluation activities are in Intensive Evaluation for Criminal Justice Planning Agencies (Wiedman, et al., 1975), Practical Program Evaluation for State and Local Government Officials (Hatry, et al., 1973), and Federal Evaluation Policy (Wholey, et al., 1976). The first of these publications also gives several examples of alternative organizational strategies used by SPAs in the past for conducting intensive evaluation.

TECHNICAL STEPS

Recommended Resources

The remaining steps in evaluation are technical. These are clearly laid out in many research texts and in government manuals for agency personnel responsible for conducting or overseeing evaluations. For example, Evaluation Research: Methods of Assessing Program Effectiveness (Weiss, 1972) addresses in general the application of research techniques to the evaluation of social programs. It formulates the questions and various research designs and then addresses the political context in which such evaluation takes place. The presentation is cogent and easy to follow. It would be useful for an agency or project manager in evaluating staff capabilities and the time and money likely to be needed. Other resources are:

- Practical Program Evaluation for State and Local Government Officials (Hatry, et al., 1973) is directed at the government analyst. It lays out an easily followed outline of considerations and steps associated with program evaluation.

- Routinizing Evaluation: Getting Feedback on Effectiveness of Crime and Delinquency Programs (Glaser, 1973) is a somewhat longer narrative. Examples provided by Glaser are more specific to crime and delinquency than in the Weiss or Hatry volumes, but the format is less straightforward and less introductory in nature. It goes beyond impact evaluation and discusses procedures associated with process evaluation.

- Federal Evaluation Policy (Wholey, et al., 1976) is a broad treatment of the need for a wide-ranging evaluation strategy for large government programs and a discussion of administrative systems, organizational relationships, responsibilities, and methodology.

- Evaluation in Criminal Justice Programs: Guidelines and Examples (Albright, et al., 1973) was prepared as a guide for LEAA National Impact Program managers. It discusses developing and implementing plans to evaluate criminal justice projects and programs and gives examples of applying evaluation methodology and of effectiveness and efficiency measures.

- Intensive Evaluation for Criminal Justice Planning Agencies (Weidman, et al.) defines, according to LEAA, "intensive evaluation" and describes ways it can be carried out. It provides examples of alternative organizational strategies for intensive evaluation used by four SPAs, along with brief examples of strategies used in states with small-sized grants and of data systems used by SPAs for evaluation purposes.

Technical Tasks

Initial tasks require a series of decisions:

- What is to be measured?
- How is it to be measured?
- Who is to be measured?
- When is it to be collected and for what groups?
- How it is to be collected?
- How are the data to be used?

The reader should not be misled into thinking that selecting a design and carrying out the research will be simple and predictable. Any of the publications listed above should make this evident. In particular, the Weiss and Glaser volumes discuss not only the techniques but also the pitfalls associated with research efforts, especially those conducted in social and political contexts for use in decision making (as compared to basic research for the sake of knowledge).

TASK 1: Formulating the Question

An evaluation cannot be designed until program objectives are clarified and stated in measurable terms. For example, "To enhance the well-being of youth" is not measurable; "To reduce truant behavior for youth in this program" lends itself better to devising a measure of whether or not this has been accomplished. If there are multiple objectives, one or two must be selected as more important in order to provide a manageable study. The staff itself and the staff and the evaluator must agree about this definition of program objectives, about the importance of one or two over others, about the target group to be measured, and about the preference of studying short-term effects, long-term effects, or both.

Similar decisions and definitions need to be made with respect to a process evaluation. There must be a clearcut rationale for having collected certain information. This underlying reason guides compilation of data and helps judge the project's efficiency and activities.

TASK 2: Designing Instruments or Techniques for Measurement

Perhaps the best strategy here is to locate suitable existing measures. For example, carefully developed and documented scales exist for many psychological variables, such as alienation or self-esteem. Handbooks are available which describe established measures (see Lake, Miles, and Earle; 1973). This approach eliminates much trial and error which

accompanies the creation of new measures. It may also provide for responses in earlier investigations which could be used for comparison purposes.

If there are no suitable measures available, then careful work must be done at this stage. Hastily designed instruments or techniques at the beginning of a study can lead to increased difficulties with making sense of the findings, or even jeopardize their utility. The theory of measurement is highly developed and much has been written on it. It would be irresponsible for a project to create measures for evaluative research without familiarity with the various types and levels of measurement. Some measurements are more precise than others; however, project design or objectives may not always allow for the most precise ones to be used. This needs to be examined and given careful consideration to avoid aiming for a false precision.

TASK 3: Designing the Study; Common Models for Evaluation Studies

The usefulness of the findings greatly depends on this stage. If precise statements about the extent to which project activities are responsible for outcomes are most important, then sophisticated and complex designs must be used, and used correctly. Other designs are less complex than classic experimental designs to execute and analyze, but these are also less precise in linking effects to activities and they leave far more room for alternative explanations for the actual outcomes.

There are many different models incorporating elements of experimental design that are appropriate to evaluative research and present a range of explanatory power and ease of implementation. For example, Federal Evaluation Policy presents a chart of eleven pre-experimental, true experimental, and quasi-experimental designs which have been developed by social scientists and discusses their relative strengths and constraints. The classic work upon which the discussion is based is Experimental and Quasi-Experimental Designs for Research (Campbell and Stanley, 1966), written for a more technical audience. Below are four common evaluation models. These are discussed in Intensive Evaluation for Criminal Justice Planning Agencies and are among those commonly described in any presentation of how to conduct program evaluation. Practical Program Evaluation for State and Local Government Officials describes these same four designs in somewhat more detail, addressing general steps involved in each one, problems likely to be encountered in each, types of application, and relative costs. These can serve as a guide for project staff considering setting up an impact evaluation. However, it is recommended that they work with someone with expertise and experience in setting up such studies because it is easy to view these designs as being simpler to implement than is actually the case.

Designs 2, 3, and 4 on the following chart allow for controlling some of the other factors that might affect the target group. Design 4, which is the classic controlled experiment design, is the most powerful

in terms of being able to attribute program effects to program activities. It is also the most difficult to prepare, carry out, and analyze and will take more planning, more expertise, and more staff time than any of the others. Design 1 is probably the most commonly used. It is the simplest and cheapest type of evaluation and can identify if there has been a change. But, at the same time, it is the least capable of separating the effect of program activities from other influences.

A fifth model of measurement (not actually a design) compares actual results against planned or targeted results. This is not an impact measurement because it does not measure the effects of service, but often can provide adequate information for planning purposes. It is generally based on such things as measures of workload and population served and can be accomplished with information collected in reporting systems designed for program monitoring.

TASK 4: Data Collection

Data collection is the most routine of the tasks, but it is likely to consume the greatest amount of time and effort. It takes thoughtfulness and commitment on the part of the staff, for if done sloppily or inconsistently the legitimacy of the findings is reduced. Data must be accurate, complete, and comparable from time period to time period or from group to group.

Though we commonly think of interviews or questionnaires as basic data collection methods, using existing records and statistics is probably more prevalent. This would be particularly true for process evaluation and is by definition true when the idea of evaluation has not been thought of until the end of the project. Evaluators with ingenuity can often think of other special data collection procedures such as systematic field observations, special event analysis, critical incident techniques, and so forth. These may make it more difficult or even impossible to obtain the desired "before" measures, but should not be overlooked if other collection methods are not applicable to the project's objectives or operations.

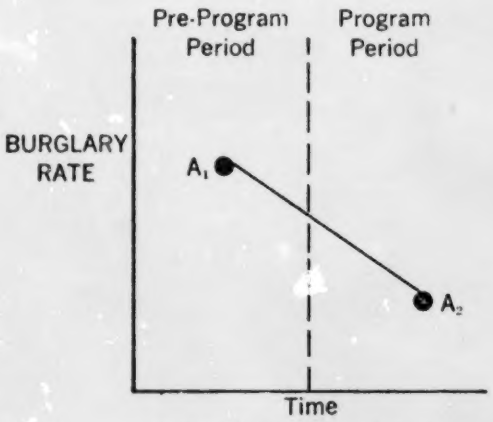
TASK 5: Utilization of Results

If the suggested steps are followed, staff already know how the results are to be used and who is going to use them, as this was a premise in deciding whether or not the evaluation should be undertaken in the first place.

The most important thing to understand at this point is that the data will not speak for themselves. The staff must be prepared to think about what they mean and to consider alternative explanations for specific findings. They should be prepared to consider the findings as one piece of information and not the only piece. They should be

DESIGN 1

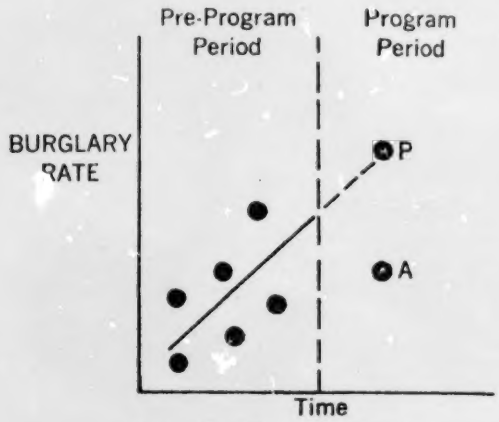
Before vs. After
Program Comparison



Estimated Program Effect = After Program Rate (A_2) – Before Program Rate (A_1)

DESIGN 2

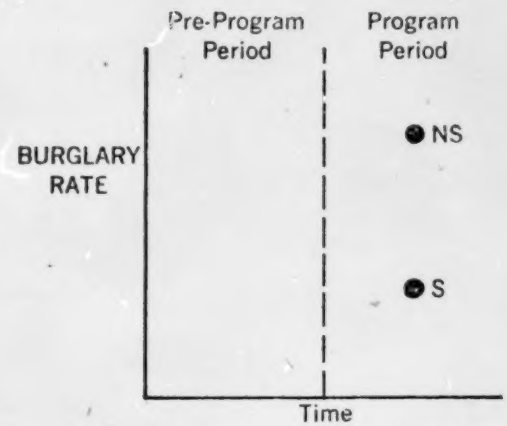
Time Trend Projection
of Pre-Program vs. Actual Program Data



Estimated Program Effect = Actual Rate (A) – Projected Rate (P)

DESIGN 3

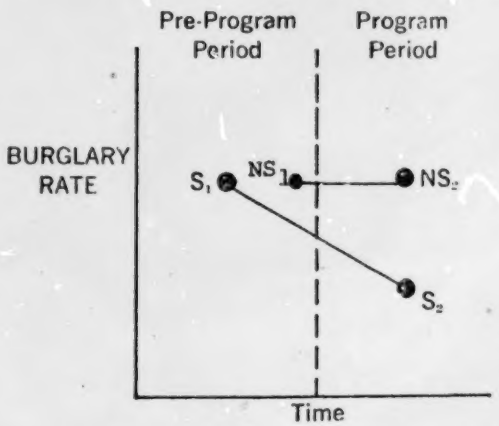
One Time Comparison Between Those
Served and Not Served by Program



Program Effect = Rate of Those Served (S) – Rate of Those Not Served (NS)

DESIGN 4

Time Series Comparisons Between
Those Served and Not Served



Program Effect = Difference in Before and After Burglary Rates of Population Served by the Program (S_2–S_1) – Similar Difference for Those Not Served (NS_2 -NS_1)

Evaluation Design Methodologies

Adapted from: Intensive Evaluation for Criminal Justice Planning Agencies
(Weidman, et al., 1975)

prepared to determine with honesty whether or not the evaluation was carried out well enough and without too many unanticipated events in the environment or in the data collection procedures. They must be prepared to ask if the product can be used as intended, if it can be used but only with qualifications, or if it should be scrapped.

Despite early agreements about the utilization to be made of the findings, the results may prove "unpopular" and resistance to using them even for formative purposes may arise. Evaluations done for summative purposes may be discounted as inaccurate, even when design and data collection can be declared adequate. Too many evaluation studies simply disappear. This is the reason for deciding to use the findings and stating what they are to be used for and for making the first decision -- not the final one.

APPENDIX A

An Administrative Checklist for Reviewing Evaluation Plans

Conceptualization of Evaluation

____Definition:	How is evaluation defined in this effort?
____Purpose:	What purpose(s) will it serve?
____Questions:	What questions will it address?
____Audiences:	Who will it serve?
____Agents:	Who will do it?
____Process:	How will they do it?
____Standards:	By what standards will their work be judged?

Sociopolitical Factors

____Involvement:	Whose sanction and support is required, and how will it be secured?
____Internal communication:	How will communication be maintained between the evaluators, the sponsors, and the system personnel?
____Internal credibility:	Will the evaluation be fair to persons inside the system?
____External credibility:	Will the evaluation be free of bias?
____Security:	What provisions will be made to maintain security of the evaluative data?
____Protocol:	What communication channels will be used by the evaluators and system personnel?
____Public relations:	How will the public be kept informed about the intents and results of the evaluation?

Contractual/Legal Arrangements

____Client/evaluator relationship:	Who is the sponsor, who is the evaluator, and how are they related to the program to be evaluated?
____Evaluation Products:	What evaluation outcomes are to be achieved?
____Delivery Schedule:	What is the schedule of evaluation services and products?
____Editing:	Who has authority for editing evaluation reports?
____Access to data:	What existing data may the evaluator use, and what new data may he obtain?
____Release of reports:	Who will release the reports and what audiences may receive them?
____Responsibility and authority:	Have the system personnel and evaluators agreed on who is to do what in the evaluation?
____Finances:	What is the schedule of payments for the evaluation, and who will provide the funds?

The Technical Design

____Objectives and variables:	What is the program designed to achieve, in what terms should it be evaluated?
____Investigatory framework:	Under what conditions will the data be gathered, e.g., experimental design, case study, survey, site review, etc?
____Instrumentation:	What data-gathering instruments and techniques will be used?
____Sampling:	What samples will be drawn, how will they be drawn?
____Data-gathering:	How will the data-gathering plan be implemented, who will gather the data?
____Data storage and retrieval:	What format, procedures, and facilities will be used to store and retrieve the data?
____Data analysis:	How will the data by analyzed?
____Reporting:	What reports and techniques will be used to disseminate the evaluation findings?
____Technical adequacy:	Will the evaluative data be reliable, valid, and objective?

The Management Plan

____Organizational mechanism:	What organizational unit will be employed, e.g., an in-house office of evaluation, a self-evaluation system, a contract with an external agency, or a consortium-supported evaluation center?
____Organizational location:	Through what channels can the evaluation influence policy formulation and administrative decision making?
____Policies and procedures:	What established and/or ad hoc policies and procedures will govern this evaluation?
____Staff:	How will the evaluation be staffed?
____Facilities:	What space, equipment, and materials will be available to support the evaluation?
____Data gathering schedule:	What instruments will be administered, to what groups, according to what schedule?
____Reporting schedule:	What reports will be provided, to what audiences, according to what schedule?
____Training:	What evaluation training will be provided to what groups and who will provide it?
____Installation of evaluation:	Will this evaluation be used to aid the system to improve and extend its internal evaluation capability?
____Budget:	What is the internal structure of the budget, how will it be monitored?

Moral/Ethical/Utility Questions

____Philosophical stance:	Will the evaluation be value free, value based, or value plural?
____Service orientation:	What social good, if any, will be served by this evaluation; whose values will be served?
____Evaluator's values:	Will the evaluator's technical standards and his values conflict with the client system's and/or sponsor's values; will the evaluator face any conflict of interest problems; and what will be done about possible conflicts?
____Judgments:	Will the evaluator judge the program; leave that up to the client; or obtain, analyze, and report the judgments of various reference groups?
____Objectivity:	How will the evaluator avoid being co-opted and maintain his objectivity?
____Prospects for utility:	Will the evaluation meet utility criteria of relevance, scope, importance, credibility, timeliness, and pervasiveness?
____Cost effectiveness:	Compared to its potential payoff, will the evaluation be carried out at a reasonable cost?

Used with the permission of Daniel L. Stufflebeam--from *Meta Evaluation*. Paper #3, Occasional Paper Series. Kalamazoo, Michigan: The Evaluation Center, Western Michigan University, 1975.

REFERENCES

ALBRIGHT, ELLEN, MARTIN A. BAUM, BRENDA FORMAN, SOL GEMS, DAVID JARRE, FRANK C. JORDAN, JR., RUTH KATZ, and PHILIP A. SINSKY. 1973 *Criminal Justice Research: Evaluation in Criminal Justice Programs, Guidelines and Examples.* US Department of Justice, Law Enforcement Assistance Administration, National Institute of Law Enforcement and Criminal Justice. Washington, DC: US Government Printing Office. (June.)

CAMPBELL, DONALD T. and JULIAN C. STANLEY. 1966 *Experimental and Quasi-Experimental Designs for Research.* Chicago: Rand McNally and Company.

GLASER, DANIEL. 1973 *Routinizing Evaluation: Getting Feedback on Effectiveness of Crime and Delinquency Programs.* Rockville, Maryland: National Institute of Mental Health, Center for Studies of Crime and Delinquency.

HATRY, HARRY P., RICHARD E. WINNIE, and DONALD M. FISK. 1973 *Practical Program Evaluation for State and Local Government Officials.* Washington, DC: The Urban Institute.

LAKE, DALE G., MATTHEW B. MILES, and RALPH B. EARLE, JR., eds. 1973 *Measuring Human Behavior: Tools for the Assessment of Social Functioning.* New York: Teachers College Press.

LAW ENFORCEMENT ASSISTANCE ADMINISTRATION. 1974 *Monitoring for Criminal Justice Planning Agencies.* (August 1.)

SCRIVEN, MICHAEL. "The Methodology of Evaluation." 1967 In *Perspectives of Curriculum Evaluation,* pp. 39-83. Edited by Ralph W. Tyler, Robert M. Gagne, and Michael Scriven. AERA Monograph Series on Curriculum Evaluation, No. 1. Chicago: Rand McNally and Company.

STUFFLEBEAM, DANIEL L. 1975 *Meta Evaluation.* Paper #3, Occasional Paper Series. Kalamazoo, Michigan: The Evaluation Center, Western Michigan University.

SUCHMAN, EDWARD A. 1967 *Evaluative Research: Principles and Practice in Public Service and Social Action Programs.* New York: Russell Sage Foundation.

WALLER, JOHN D., DONA MACNEIL KEMP, JOHN W. SCANLON, FRANCINE TOLSON, and JOSEPH S. WHOLEY. 1976 *Monitoring for Government Agencies.* Washington, DC: The Urban Institute. (February.)

WEIDMAN, DONALD R., JOHN D. WALLER, DONA MACNEIL, FRANCINE L. TOLSON,
1975 and JOSEPH S. WHOLEY. *Intensive Evaluation for Criminal Justice Planning Agencies*. US Department of Justice, Law Enforcement Assistance Administration, National Institute of Law Enforcement and Criminal Justice. Washington, DC: US Government Printing Office. (July.)

WEISS, CAROL H. *Evaluation Research: Methods of Assessing Program*
1972 *Effectiveness*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

WHOLEY, JOSEPH S., JOHN W. SCANLON, HUGH G. DUFFY, JAMES S. FUKUMOTO,
1976 and LEONA M. VOGT. *Federal Evaluation Policy: Analyzing the Effects of Public Programs*. Washington, DC: The Urban Institute.

☆U.S. GOVERNMENT PRINTING OFFICE: 1979 O— 281-387/21